DOI:10.11931/guihaia.gxzw201803032

# 植物 DNA 条形码研究领域文献计量学及可视化分析*

熊 勇[1,2]，李文义[1]， 杨淬[1]，罗斌圣[2]，杨青松[1]

（1.云南民族大学民族医药学院，昆明 650500; 2.中央民族大学生命与环境科学学院，北京 100081）

**摘 要：**为了全面了解植物 DNA 条形码研究领域的发展和最新动态，探讨中国 DNA 条形码发展的状态和前景。利用 Web of Science 数据库对该研究领域进行文献计量学统计，并对引用频次、研究热点和研究前沿进行了可视化分析。获得了中国、美国、加拿大学者在该领域文献贡献率最大，中国研究机构发文量领先，但美国、加拿大科研机构论文质量较高，影响力较大。文献引用频次可视化发现，2009 年是该领域研究的高峰期，该研究领域的前沿和研究热点主要集中在物种的识别和生物多样性应用、DNA 条形码候选序列筛选和鉴定技术的规范化。中国学者在植物 DNA 条形码领域研究具有领军作用和很高的影响力，国家提倡中药产业的发展也推动了我国 DNA 条形码蓬勃发展，但是论文的质量和影响力与美国、英国、加拿大等发达国家研究还有一定的差距，应该加大与发达国家科研机构合作，提高研究能力，DNA 条形码技术在植物的鉴定、分类和生物多样性的保护起到非常重要作用。建立一个更全面、通用的全球植物 DNA 条码库和开发新的标记并采用新的测序技术是植物 DNA 条形码研究的未来前景。

**关键词：**植物 DNA 条形码，文献计量学，生物多样性，citespace，可视化分析

# Bibliometric and visualization analysis of DNA barcoding in plants

XIONG Yong[1, 2], LI Wenyi[1], YANG Cui[1],
LUO Binsheng[2], YANG Qingsong[1]

(1. *School of Ethnic Medicine, Yunnan Minzu University*, Kunming 650500, China；2. *College of Life and Environmental Sciences, Minzu University of China*, Beijing 100081, China)

**Abstract:** In order to get the latest information and development researches filed of DNA barcoding of plants, and discuss its role in biodiversity conservation. Used bibliometrics, Histcite and citespaces to analyze literatures of DNA barcoding in plants based on Web of Sciences database to, and analyze Citation counts, research hot and research front visualization. The results showed the rate of contribution of literatures was the biggest of Chinese, American, Canadian authors. Chinese institutions published literatures counts than others countries, while literatures from American and Canadian institutions have more higher quality and influence. This research field has gotten peak at 2009 based on citation visualization analysis. The research front and hot focus on species identify and biodiversity application, find the suited candidates DNA barcoding sequences and refining the identical technology. Chinese authors play a leading role and have strength influence at DNA barcoding of plants. Chinese government encourages traditional

**基金项目：**国家自然科学基金（81760655）；云南省应用基础研究计划项目（2016FD057）[Supported by the National Natural Science Foundation of China(81760655)；the Program Applied Basic Research Foundation of Yunnan Province(2016FD057)]

**作者简介：**熊 勇(1977-)，男，云南罗平人，副教授，在读博士研究生，主要从事民族植物学研究，(E-mail)17639121@qq.com.

Chinese medicine industry to promote the development of DNA barcoding technology at China. But the quality and influence of published literatures have a gap between China and developed countries, Chinese institutions should increase cooperation with the developed countries institutions and promote itself research ability. It is very important that DNA barcoding technology at species identify and biodiversity conversation. Tomorrow's Outlook for Plant DNA Barcoding is building the global plant DNA barcode library and new DNA markers and new sequencing technologies.

**Key words:** DNA barcoding of plants, bibliometric, biodiversity, citespace, visualization analysis

DNA 条形码技术是利用标准的基因片段对物种进行快速鉴定(Hebert et al, 2003), 该技术提供了可信息化的分类学标准和有效的分类学手段，已经被成功用于生物物种鉴定和分类(Liu et al, 2011)，生物多样性调查(Lahaye, 2008)和生态学研究(Valentini, 2009)等领域，并成为进展最迅速的学科前沿之一。生命条形码数据 (BOLD) 系统提供了主要针对动物类群 DNA 条形码研究的技术规范，由于植物本身的生物学特性与所使用的条形码不同，植物 DNA 条形码研究相对滞后些，然而其作为一个 DNA 条形码的研究领域重要内容之一，植物 DNA 条形码不仅受到植物学研究的影响，又同时受到了分子生物学和生物信息学相互渗透和影响。DNA 条形码在植物中研究及应用不断增多， 2004 年 Blaxter ML 对 DNA 条形码的在分类中应用前景进行了分析(Blaxter,2004)，2005 年 Kress WJ 等利用 DNA 条形码技术鉴定开花植物(Kress et al, 2005)，2007 年 Chase MW 提出关于陆生植物条形码的标准化方法的建议(Chase et al, 2007)，2009 年 CBOL Plant Working Group 开展 DNA 条形码技术在陆生植物的应用，2017 年 Bashir Mohammed Abubakar 对 DNA 条形码在草药产品中的鉴定应用进行回顾和总结(Mohammed et al, 2017)，DNA 条形码技术在植物上的研究和应用在源源不断的增加。

文献计量学是利用数学和统计学方法来分析相关知识载体，从而获取有价值信息的一门交叉学科(赵蓉英, 2010)。一个学科或一个研究领域的发展情况，必然表现在相关论文的发表和被引用方面，因此文献计量学可以用于分析、描述学科发展状况和预测学科发展趋势，再通过绘制知识图谱将结果以图形的形式直观、形象地表达出来(顾洪涛, 2013)，目前用文献计量学对 DNA 条形码在植物上的研究领域总体分析见报道的不多，然而国内外已经有很多基于 Web of Science 数据库对文献进行文献计量学统计分析，例如孙秀焕等基于 Web of Science 分析了水稻研究态势(孙秀焕, 2012)，杨华等进行了国际茶多酚类研究文献发展态势研究(杨华等，2013)，Garfield 利用 HistCiter 软件对科学史进行可视化分析（Garfield，2009）。本研究基于 Web of Science 数据库，对 2003~2016 年植物 DNA 条形码国际文献进行文献计量学统计和分析，再利用 Histcite 和 Citespace 软件对该研究领域进行了可视化分析，达到对植物 DNA 条形码研究领域文献信息全面了解。

## 1.数据来源与数据处理

### 1.1 数据来源

以 Web of Science（WoS）核心合集为数据来源，时间跨度为 2003—2016 年，以(“DNA barcoding” or “DNA bar coding” or “DNA barcode” or “DNA barcodes”) plant 为 Topic 进行精确检索，下载文献信息保存为纯文本文档，为后续分析做准备。

### 1.2 数据处理

#### 1.2.1 Web of Science 文献处理

通过在线 http://ip-science.thomsonreuters.com/thanks/histcite/ 网站，获得 Histcite 软件，同时把 Web of Science 获得的文献信息整合为一个完整的 txt 文档，导入 Histcite 软件，基于 Web of Science 创建引文报告和分析检索结果以及 Histcite 统计，分析的内容主要包括：每

年出版的文献量、被引频次、核心作者、出版的期刊、主要研究机构和国家/地区等（Shadbolt et al, 2013; Bharathi, 2013）。

1.2.2 Histcite 文献处理

以 LCS 作为节点，再用其中的 Graph maker 制作引文编年图，探究文献间的引用关系及获得重要文献。

1.2.3 研究热点的可视化分析

将分析后获得的文献信息导入 Citespace 软件中，对该领域研究热点、研究前沿和突现点进行可视化分析。

## 2. 结果与分析

用植物 DNA 条形码为主题在 Web of Science 核心合集进行检索，共获得检索结果 1 097 条，这些论文被引频次总计 27 125 次，去除自引 18 922。

**2.1 发文量和引文量分析**

基于 Web of Science 核心合集，时间段位 2003—2016 年，以 DNA 条形码为关键词进行检索，用 Web of Science 在线工具进行引文报告分析，通过创建引文报告获得每年文献量和被引频次（图 1 和图 2）。

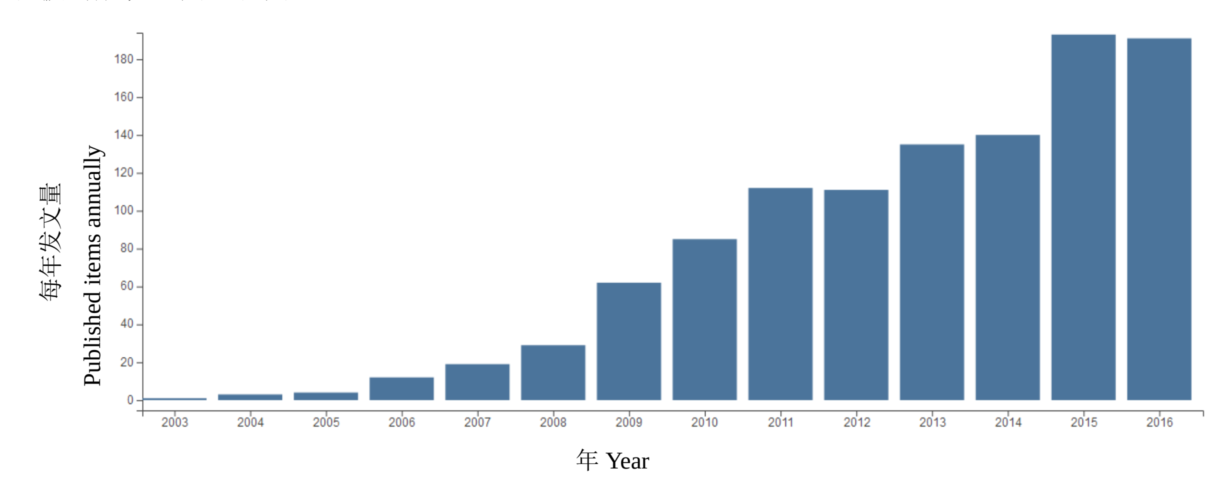



图 1 每年出版的文献数

Fig.1 Published items annually

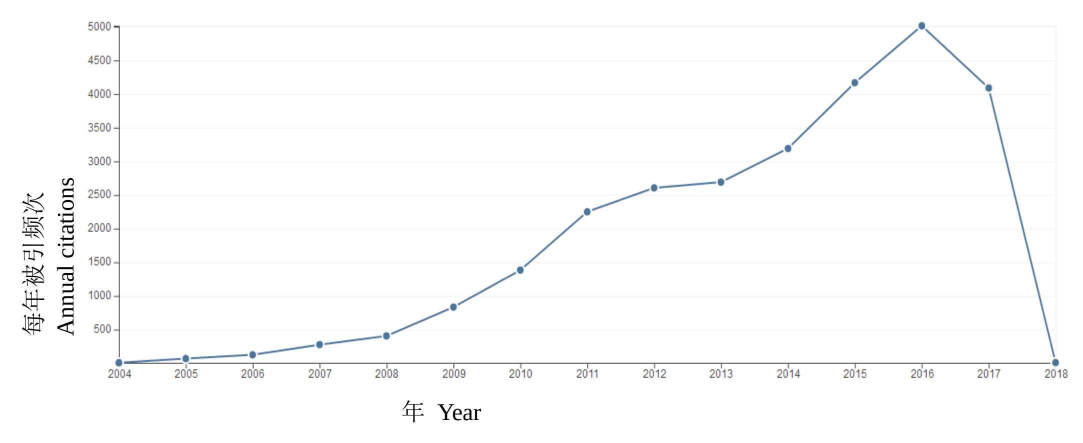



图 2 按年份的被引频次

Fig.2 Citation frequencies annually

从图 1 和图 2 可以看出每年文献量从 2003 年的个位数篇到 2016 接近 200 多篇，文献发文量不断增多，每年的被引频次从 2004 年 0 次到 2016 年接近 5 000 次，增加的速率不断加大，从 2009 年后文献量和论文被引用数量，呈现快速的增长，表明该研究领域稳定发展。

**2.2 论文被引数量分析**

对获得原始文档的适当修改，导入 Histcite 软件，进行 LCS(Local Citation Score )和 GCS（Global Citation Score）指标分析（图 3，图 4）。LCS 是指论文在本地数据集（WoS 中输入关键词后导出的所有文献）中被引用次数，LGS 值越高表示该论文在该研究领域内的重要性高，GCS 值指论文在整个 WoS 数据中的总被引次数，但施引论文不一定是该领域内的论文。

图 3 和图 4 为 2003 年至 2016 每年 TLCS(Total Local Citation Score)和 TGCS(Total Global Citation Score）的统计图。从图中可以看出，TLCS 和 TGCS 都在 2007 年至 2012 年这几段达到较高水平，而后半段 2013 年开始都呈下降趋势，这种趋势与每年的发文量和论文引用数量不一致。TLCS 和 TGCS 在 2009 年都达到了高峰期，2009 年发的文献对该领域研究具有重要的作用和很大的影响力，其中 CBOL Plant Working Group、Hollingsworth ML 等合作发表的论文《A DNA barcode for land plants》和《Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants》（Hollingsworth et al,2009)，前者对现有的 7 条候选 DNA 条形码进行对比，推荐 rbcL + matK 组合作为陆生植物 DNA 条形码的标准，后者是对陆生植物 DNA 条形码提出了 7 条候选序列（atpF-atpH spacer, matK gene, rbcL gene, rpoB gene, rpoC1 gene, psbK-psbI spacer, and trnH-psbA spacer），两篇论文对后面陆生植物 DNA 条形码的研究起到了指导性作用。从后面的文献被引用频次的可视化分析得出，依据 TLS 值获得排名前 30 的论文，其中 2009 年附近文献就有 13 篇，可见 2009 年是植物 DNA 条形码的研究的高峰期和热点区。

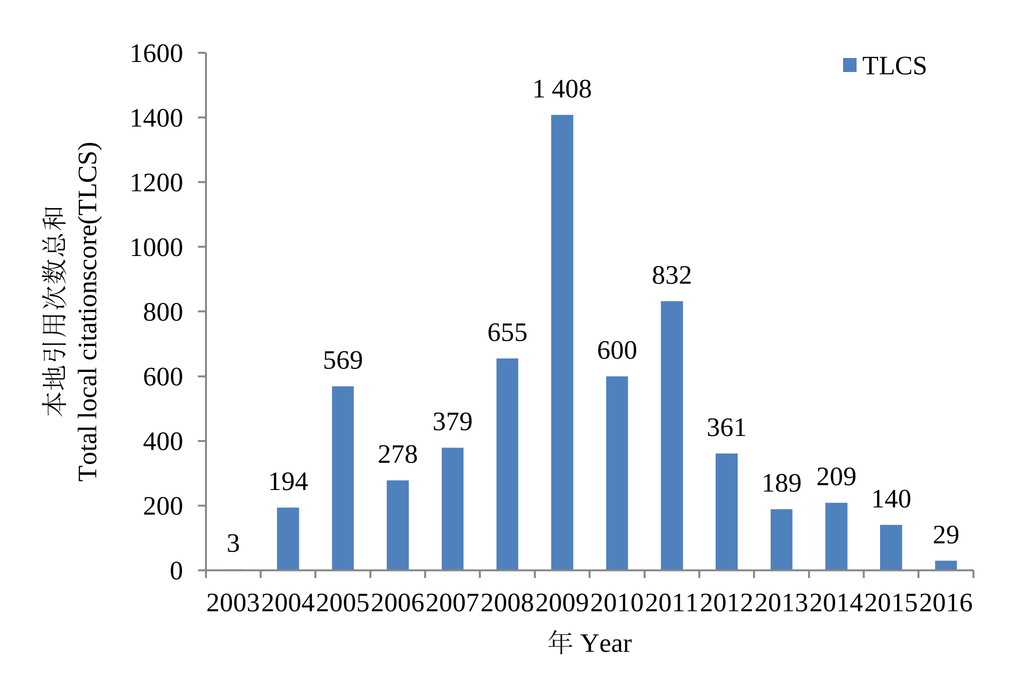



图 3 每年论文被引数统计（TLCS）

Fig. 3 Annual document cited statistics（TLCS）

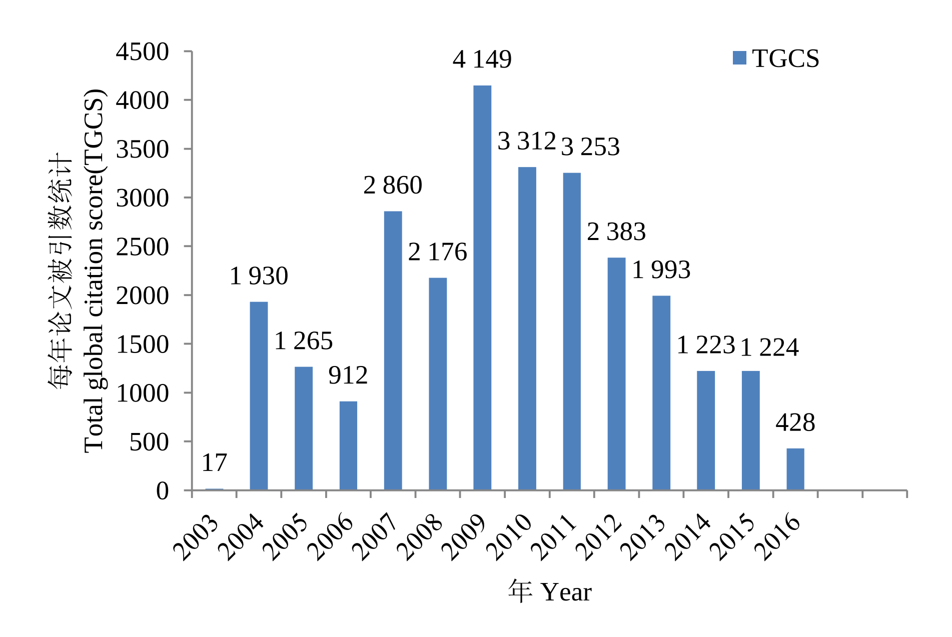



图 4 每年论文被引数统计（TGCS）

Fig. 4 Annual document cited statistics（TGCS）



### 2.3 核心作者分析

把获得文献信息导入 Histcite 软件进行分析，获得 2003—2016 年发表有关植物 DNA 条形码英文论文的作者有 4 023 位，筛选出发文量前 10 的作者（表 1），其中有 5 名为中国学者，说明中国在这一领域研究文献贡献率很大，但是从文献的被引用频次上和影响力上，与国外学者有一定的差距，核心作者中是美国 Smithsonian Institution 的 Kress WJ 的 TLCS 最高，其次是加拿大 University of Guelph 的 Newmaster SG，中国中医科学院中药研究所陈士林教授在中药材 DNA 条形码研究具有很高的成就，所编写《中国药典中药材 DNA 条形码标准序列》是该行业指导性书籍，发文量第二是中国医学科学院药用植物研究所的宋经元教授，在中药 DNA 条形码鉴定研究中起到领头羊的作用。

表 1 论文数量前 10 名学者

Table 1 Top 10 authors with most publications

| 序号 No. | 作者 Author | 发文数（篇） Recs | TLCS | TGCS |
|---|---|---|---|---|
| 1 | Chen SL | 48 | 634 | 1 936 |
| 2 | Song JY | 32 | 390 | 1 463 |
| 3 | Newmaster SG | 29 | 907 | 1 928 |
| 4 | Yao H | 29 | 305 | 1 241 |
| 5 | Kress WJ | 25 | 1 013 | 2 723 |
| 6 | Hebert PDN | 21 | 259 | 2 220 |
| 7 | Ragupathy S | 18 | 264 | 549 |
| 8 | Hajibabaei M | 17 | 568 | 1 488 |
| 9 | Li DZ | 17 | 273 | 571 |
| 10 | Pang XH | 17 | 230 | 1 085 |

### 2.4 国家（地区）和研究机构分析

对国家（地区）和研究机构文献量进行统计，排名前 10 见表 2 和表 3，从表 2 可以得出，发文量前五的国家都是在 DNA 条形码研究领域进行比较早或者具有植物研究传统国家，美

国和中国发文量最多，而且数量相差不大，与其他国家（地区）相比优势很明显，但是从文献的影响力和被引用频次来看，中国与美国、加拿大和英国都有差距，可能跟我们国家科技发展相对滞后或文化有关，但是从前 10 名国家（地区）上看，该领域的研究主要集中在科技相对发达的国家和地区，中国对该领域的研究也有很多成绩和贡献。

表 2  发表文献量排名前 10 的国家和地区

Table 2 Top 10 Countries/areas based on publications

| 序号<br>No. | 国家<br>Country | 发文数（篇）<br>Recs | TLCS | TGCS |
|---|---|---|---|---|
| 1 | USA | 247 | 2 260 | 10 608 |
| 2 | Peoples R China | 245 | 1 097 | 3 569 |
| 3 | Canada | 141 | 1 395 | 6 403 |
| 4 | UK | 106 | 1 902 | 5 942 |
| 5 | India | 83 | 140 | 746 |
| 6 | Germany | 66 | 151 | 1 752 |
| 7 | Italy | 64 | 247 | 1 466 |
| 8 | France | 60 | 437 | 2 857 |
| 9 | Australia | 42 | 68 | 862 |
| 10 | Netherlands | 41 | 76 | 1 475 |

表 3  发文量前 10 位的机构

Table 3 Top 10 institutions based on publications

| 序号<br>NO. | 机构<br>Institution | 发文数（篇）<br>Recs | TLCS | TGCS |
|---|---|---|---|---|
| 1 | Chinese Academy of Sciences | 105 | 489 | 1 458 |
| 2 | University of Guelph | 69 | 1 218 | 4 390 |
| 3 | Chinese Academy of Medical Sciences | 61 | 491 | 1 233 |
| 4 | Smithsonian Institution | 45 | 1 078 | 3 168 |
| 5 | Peking Union Medical College | 28 | 117 | 723 |
| 6 | China Academy Chinese Medical Sciences | 25 | 70 | 329 |
| 7 | Royal Botanic Gardens | 24 | 1 429 | 3 052 |
| 8 | Agriculture and Agri-Food Canada | 21 | 115 | 990 |
| 9 | Natural History Museum | 20 | 575 | 1 512 |
| 10 | University of Oslo | 18 | 29 | 981 |

从表 3 的研究机构前十名中，可以得出与核心作者和国家相似的结果，进行植物 DNA 条形码研究主要机构还是集中在三个国家，中国、加拿大、美国，中国科学院发文量最多。但是以 TLCS 为指标进行排序（表 4），前三名为英国 Royal Bot Gardens，加拿大的 University of Guelph 和美国的 Smithsonian Institution，前十名研究机构没有中国研究机构。因此，从发文量来看，前十名中有四个中国研究机构，说明我国有很多学者在进行植物 DNA 研究，并且也获得不错的成果。

表 4 TLCS 前 10 位的机构

Table 4 Top 10 institutions based on TLCS

| 序号<br>No. | 机构<br>Institution | 发文数（篇）<br>Recs | TLCS | TGCS |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 1 | Royal Botanic Gardens | 24 | 1 429 | 3 052 |
| 2 | University of Guelph | 69 | 1 218 | 4 390 |
| 3 | Smithsonian Institution | 45 | 1 078 | 3 168 |
| 4 | New York Botanic Garden | 17 | 702 | 1 825 |
| 5 | University of Johannesburg | 14 | 616 | 1 132 |
| 6 | University Costa Rica | 4 | 606 | 1 117 |
| 7 | University of Penn | 18 | 577 | 2 665 |
| 8 | Natural History Museum | 20 | 575 | 1 512 |
| 9 | Universidad Nacional Autonoma de Mexico | 9 | 547 | 1 004 |
| 10 | University of Toronto | 11 | 546 | 1 450 |

**2.5 核心期刊**

对于学术期刊的发文量与引文量统计不仅可以让我们知道不同期刊在该领域的影响力，也能为学者们选择目标期刊投稿与参考文献研究提供有价值的参考依据。对相关的期刊进行分析，总共有 338 个期刊发表于植物 DNA 条形码相关论文，以发文量进行排名，前 10 名期刊见表 5，发表在 *PLOS ONE* 期刊上的文献量最大，其次是 *MOLECULAR ECOLOGY RESOURCES* 和 *GENOME*，核心期刊中大部分都是与分子和系统进化相关，*PLOS ONE* 期刊 TLCS 值为 0，而 TGCS 值比较大，说明相对而言该期刊发表的植物 DNA 条形码领域杂志的影响力较低，但是对于其他领域研究发展具有重要参考价值。

表 5 发文量前 10 的核心期刊

Table 5 Top 10 of plant DNA barcoding based on records

| 序号<br>No. | 期刊<br>Journal | 论文数（篇）<br>Recs | TLCS | TGCS |
|---|---|---|---|---|
| 1 | PLOS ONE | 116 | 0 | 3 995 |
| 2 | MOLECULAR ECOLOGY RESOURCES | 71 | 1 002 | 2 528 |
| 3 | GENOME | 37 | 12 | 116 |
| 4 | ZOOKEYS | 26 | 28 | 127 |
| 5 | GENETICS AND MOLECULAR RESEARCH | 25 | 25 | 80 |
| 6 | JOURNAL OF SYSTEMATICS AND EVOLUTION | 24 | 248 | 400 |
| 7 | MOLECULAR ECOLOGY | 20 | 55 | 456 |
| 8 | PLANTA MEDICA | 20 | 185 | 380 |
| 9 | ZOOTAXA | 20 | 10 | 84 |
| 10 | TAXON | 17 | 416 | 658 |

表 6 植物 DNA 条形码相关论文被引用数前 10 的期刊

Table 6 Top 10 of plant DNA barcoding based on TLCS

| 序号<br>No. | 期刊<br>Journal | 论文数（篇）<br>Recs | TLCS | TGCS |
|---|---|---|---|---|
| 1 | PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA | 10 | 1 549 | 4 335 |
| 2 | MOLECULAR ECOLOGY RESOURCES | 71 | 1 002 | 2 528 |
| 3 | TAXON | 17 | 416 | 658 |
| 4 | JOURNAL OF SYSTEMATICS AND EVOLUTION | 24 | 248 | 400 |
| 5 | PLANTA MEDICA | 20 | 185 | 380 |

| | | | | |
|---|---|---|---|---|
| 6 | PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES | 8 | 182 | 327 |
| 7 | CANADIAN JOURNAL OF BOTANY-REVUE CANADIENNE DE BOTANIQUE | 2 | 152 | 239 |
| 8 | JOURNAL OF ETHNOPHARMACOLOGY | 8 | 152 | 279 |
| 9 | AMERICAN JOURNAL OF BOTANY | 13 | 141 | 1 331 |
| 10 | TRENDS IN ECOLOGY & EVOLUTION | 4 | 125 | 365 |

**2.6 被引频次的可视化分析**

用 HistCite 软件中 make graph 功能，以 LGS Count 为条件，设节点为 30，绘制出植物 DNA 条形码引文编年图（图 5）。

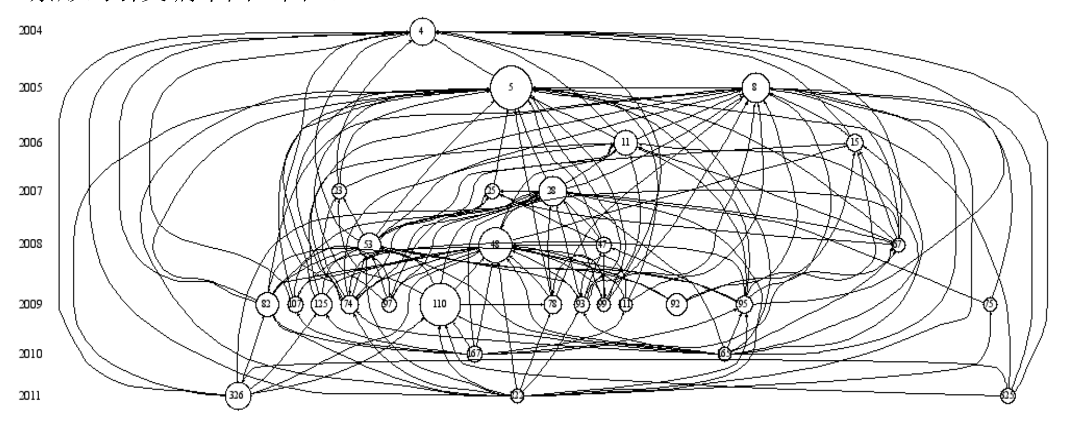



图 5 文献被引用频次的可视化分析

Figure 5 Visual analyses of ethnoveterinary literatures based on citations

如图 5 所示，从上到下是年份（2004—2011），箭头指向的文献是被引用的文献，方框内所标数字是该节点文献在所分析文献中的序号，文献引用频次越高方框越大，就是该领域比较重要的文章。图 5 展示植物 DNA 条形码研究不同文献之间引证的关系链，文献之间关系紧密，其中序号为 4，5，8，28，48，110 和 326 是被引用频次较多的文献，这些文献对该领域的研究具有重要作用。文献序号 5 和序号 110 方框最大，被引用的频次最多，文献序号 5 是 Kress et al 利用 DNA 条形码技术鉴定开花植物，该论文是较早使用 DNA 条形码技术应用于植物研究，序号 110 是 CBOL Plant Working Group 合作所写关于 DNA 条形码技术规范具有指导性的一篇论文。从年份我们看出，2009 年文献 TLS 排名前 30 最多一年，是研究的热点和高峰期，2009 年对于植物 DNA 条形码候选序列是一个寻找和验证的时期，对寻找适合的 DNA 条形码起到了奠定作用，而到了 2011 年对植物 DNA 条形码具有很大发展，多了 ITS 序列（序号 326）（Li et al, 2011）和对现在候选序列的改善（文献序号 325）（Hollingsworth, 2011）。

同时，我们也可看出，被引频次排在前 3 位的均是发达国家研究机构，可见 DNA 条形码研究的高质量和影响力大的论文还是集中在发达的西方国家研究机构。植物 DNA 条形码候选序列文章一经发表，就引起了相关学者的广泛关注，说明 DNA 条形码技术的规范化和对候选序列的优选是最前沿课题。

**2.7 热点分析**

对于一个学科的范式可视化研究主要可分为基于文献的共引分析（Co-citation）和基于词或词组的共词分析（Co-word），相较之下共引分析更广泛应用于系统揭示学科结构、学科的发展规律和发展趋势（伍若梅，2010）。植物 DNA 条形码研究领域是一个交叉学科，学科环境内不同子领域之间较难有共现的高频关键词，共词法难以反映领域间知识的输入输出。综合考虑，本研究选择共引法来分析植物 DNA 条形码的学科发展的热点与前沿。利用 Citespace 软件，对所得文献数据进行共引分析，时区选择（Timespan）为 2003—2016 年，

时间跨度（Slice Length）为 1 a，节点类型选择"Cited reference"，而筛选阈值 Thresholds（c；cc；ccv）（c 为节点出现频次、cc 为共同出现频次、ccv 为共现率）被设定为（2,2,5；3,2,10；3,3,15）。得出图 6 和图 7 两种形式的聚类分析图，而每个聚类代表着该领域的一个研究前沿。经过参数筛选，图谱中共出现了 255 个点（每个点代表一篇文献），441 条连线，其模块值（Modularity Q）为 0.7 108，轮廓值（Mean Silhouette）为 0.5 633，表明形成的聚类结构显著、置信度高且有意义。此外，采用 LLR 算法（对数似然率算法）从施引文献的关键词（K，Keyword list）中提取名词性术语对聚类进行命名（Cluster Labeling），命名结果统计如表 7。

从图 6 中可以得出，出现大的 13 个关键词：DNA barcoding、DNA barcode、land plant、identification、sequence、taxonomy、diversity、species identification、evolution、phylogeny、rbcl、plant 和 region, 从而把植物 DNA 条形码研究领域可以分为 13 个方面的聚类研究前沿，关键词字体越大表示研究的文献越多，从中我们得出，DNA 条形码在物种的鉴定、分类、进化、系统发生、生物多样性和陆生植物的应用等具有重要的作用，也是该领域研究的前沿和热点。

表 7 中显示不同的聚类，Cluster ID 为聚类后的编号，聚类的规模越大（也就是聚类中包含的成员数量越多），则编号越小，Size 代表的是聚类中所含有的成员数量，Silhouette 为衡量真个聚类成员同质性性指标，该数值越大，则代表该聚类成员的相似性越高，Mean year 代表的事该聚类中文献的平均年份，判断聚类中引用文献的远近，表 7 只列出聚类 LLR(likelihood Rate)前 10 名。

综合分析图 6 和图 7 以及表 7，同时参考聚类中的施引文献，我们可以大致了解到每个聚类所代表的的研究前沿。表 7 内一共总结了前 10 大植物 DNA 条形码的研究前沿，主要的几个聚类如下：

聚类 0 主要关注的是用植物 DNA 条形码的方法对一些难于鉴定的样品进行鉴定，如对食草动物(家畜和野生动物)的排泄来鉴定动物的饮食成分，代表论文《Universal DNA-Based Methods for Assessing the Diet of Grazing Livestock and Wildlife from Feces》（Pegard et al, 2009）。聚类 1 主要关注的是核糖体 DNA（ITS 序列），在植物多样性上的应用，《Confirming the Genetic Identity of Dendrobium fimbriatum Using an Amplification Refractory Mutation System (ARMS)》（Lu et al, 2010）。聚类 2 利用 DNA 条形码对不同物种进行快速的鉴定方法，如在真菌鉴定中的应用，《Current state and perspectives of fungal DNA barcoding and rapid identification procedures》(Begerow et al, 2010)。聚类 3 DNA 条形码在不同地区药用植物的应用，如在土耳其山区对当地药用植物的应用，《DNA barcoding study on sideritis trojana bornm. an endemic medicinal plant of ida mountain, turkey》(Tezcan, 2010)。聚类 4 主要关注 DNA 条形码在植物进化、生物信息学和资源上的应用，主要的论文《DNA barcoding in plants: evolution and applications of in silico approaches and resources》(Bhargava, 2013)。聚类 5 主要关注的是 DNA 条形码在各个科分类鉴定中的应用，如《DNA barcoding of Orchidaceae in Korea》(Kim et al, 2014)。聚类 6 利用核 DNA 和叶绿体 DNA 的方法对物种进行鉴定，《Species identification of Alnus(Betulaceae) using nrDNA and cpDNA genetic markers》(Ren, 2010)。

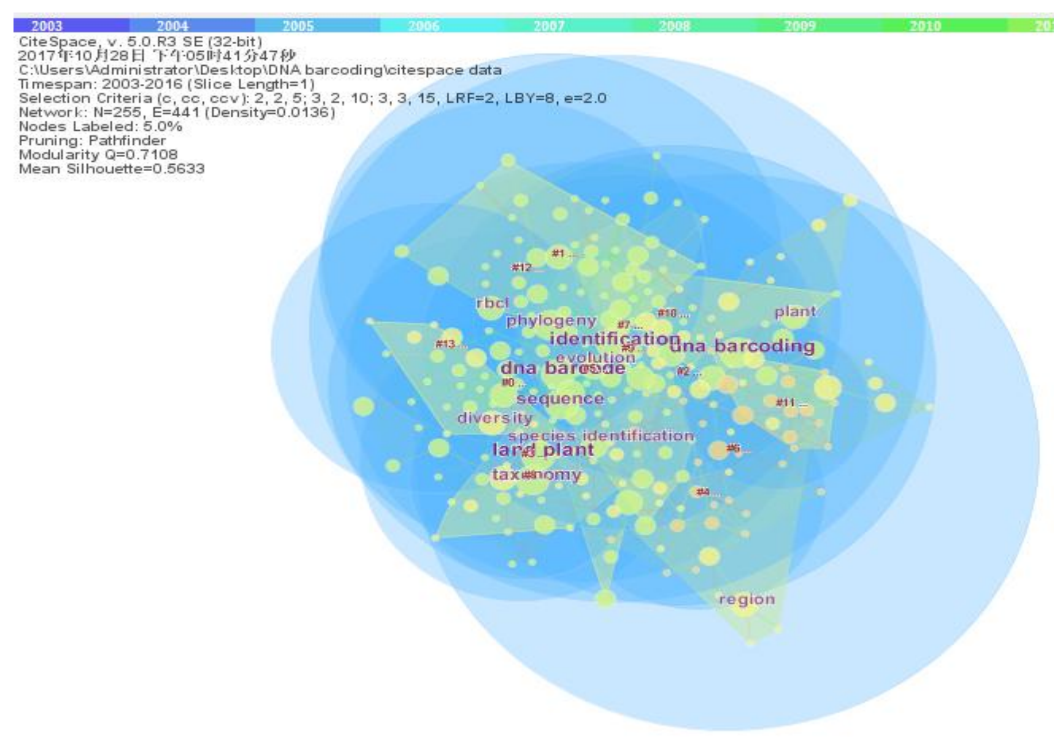



图 6 2003—2016 年植物 DNA 条形码研究前沿（Cluster view）

Fig.6 Research front DNA barcoding in plants from 2003 to 2016(Cluster view)

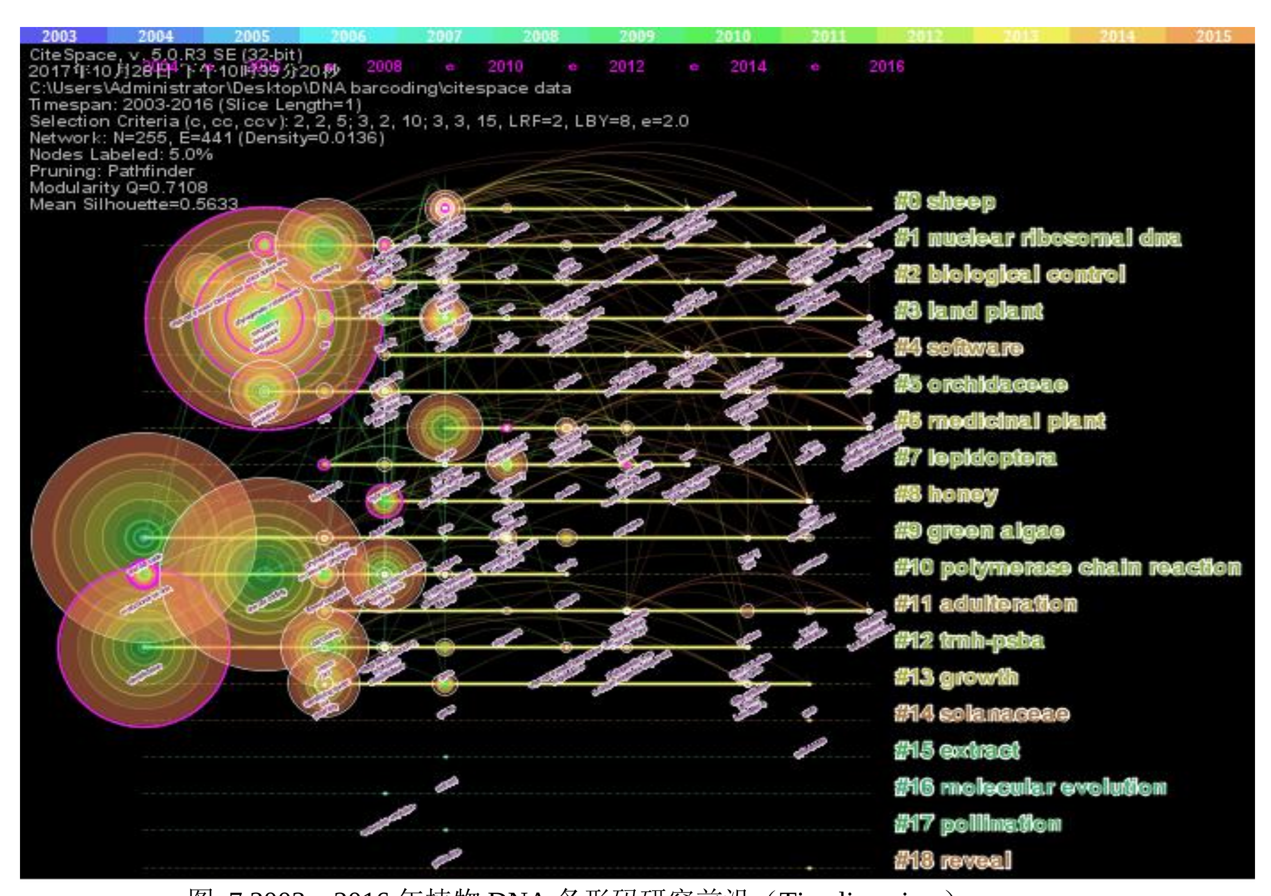



图 7 2003—2016 年植物 DNA 条形码研究前沿（Timeline view）

Fig.7 Research front of DNA barcoding in plants from 2003 to 2016(Timeline view)

表 7 前 10 个的 LLR 标签
Table 7 Top 10 of LLR

| 聚类号 Cluster ID | 大小 Size | 轮廓值 Silhouette | 平均年 Year (mean) | 对数似然率算法 LLR |
|---|---|---|---|---|
| 0 | 30 | 0.776 | 2010 | Sheep, diet analysis, polymerase, pyrosequencing, herbivore, herbivory, botanical composition, grass, conservation |
| 1 | 23 | 0.809 | 2011 | nuclear ribosomal dna, united states, phylogeny, sequence analysis, psba trnh, long distance dispersal, seed, apiaceae, gene flow, polymorphism |
| 2 | 22 | 0.695 | 2011 | biological control , parasitoid, resprouter ,,south africa, esolution melting analysis , infection phylogenetic diversity, biodiversity hotspot |
| 3 | 21 | 0.787 | 2010 | land plant, taxonomy, identify, specimen ,trnh psba, sample, sideritis trojana, life, temperature, psba trnh intergenic region |
| 4 | 18 | 0.619 | 2013 | software ,ribosomal rna ,bioinformatics , tool,database,food web,microbial community , resistance,population structure ,land plant |
| 5 | 18 | 0.596 | 2011 | orchidaceae, genome sequence ,araceae , gene sequence, root rot, size homoplasy, nucleotide sequence |
| 6 | 18 | 0.786 | 2013 | medicinal plant, region, leave, quality control, qinghai tibetan plateau,plant dna barcoding, variability, spacer, phylogenetic inference |
| 7 | 17 | 0.709 | 2010 | Lepidoptera, cryptic diversity, tachinidae , moth, parasitoid flies diptera, population genetic structure, glacial refugia, genetics , high mountain |
| 8 | 17 | 0.767 | 2013 | honey, diet, bee, species richness, system, community ecology, biodiversity, dry grassland, calcareous grassland, food traceability |
| 9 | 15 | 0.901 | 2010 | green algae, tufa, molecular phylogeny, nucleotide sequence data, series gymnobasalia, secondary structure, chlorophyta |
| 10 | 15 | 0.814 | 2007 | polymerase chain reaction, mitochondrial dna, dna taxonomy, sequence data, community structure, international species databank, bold, |

从图 8 中可以得出，2003—2016 年这段时间植物 DNA 条形码在某一时段的研究突破点或者转折点，突现的引文节点用红色表示。从图中可以，23 个主要的关键词中得出最大的节点是 2008—2011 年的鉴定（identify）关键词，最小为 2007 年至 2009 年的生命（life）关键词。DNA 条形码利用标准的基因片段对物种进行快速鉴定是研究的前沿，也是 DNA 条形码的主要作用和价值，结合前面的每年论文被引数统计和被引频次的可视化分析，可知 2008~2011 是 DNA 条形码领域研究的高峰期和热点区，对该领域的具有巨大的影响和贡献。

## 3.讨论

本研究通过 Web of Science 数据库对 DNA 条形码进行检索，基于文献计量学方法，从每年出版的文献量、被引频次、核心作者、出版的期刊、主要研究机构和国家/地区、研究热点进行了较全面的计量分析，并对被引频次、研究前沿进行了可视化分析。

（1）在国家/地区分布方面，美国、中国、加拿大论文产出数量最多，但是从文献的影响力和被引用频次来看，中国与美国、加拿大和英国都有一定的差距。（2）在研究机构方面，优秀的水稻研究机构集中在中国和美国，在发文量前三名有机构有两个是中国研究机构，但在影响力上（TLCS），中国的研究机构没有排在前十名，中国科研机构在提高自身论文水平的同时，还应注意加强与美国和其他国家优秀科研机构的交流与合作，提高自身论文水平。（3）在论文作者方面，发文量、总被引频次，主要集中在美国、中国、英国和加拿大，发

文量前 10 的作者（表 1），其中有 5 名为中国学者，说明中国在这一领域研究文献贡献率很大，但是影响力与欧美国家的学者，还有一定的距离。（4）在论文期刊分布上，优秀期刊主要集中在美国、英国、加拿大等欧美国家，没有中国的英文期刊，通过对被引频次的可视化分析，可以看出 DNA 条形码候选序列筛选和技术的应用是该领域最大关注点。（5）DNA 条形码研究热点集中 DNA 条形码在陆生植物、物种鉴定、分类、生物多样性、候选序列筛选、系统进化上的应用等方面。




图 8 2003—2016 年植物 DNA 条形码不同时期的研究突现点

Fig. 8 Citation burst of DNA barcoding in plants from 2003 to 2016

近 10 多年来，中国 DNA 条形码研究论文数量一直处于上升态势，发文量居世界第二，但论文影响力有限。中国科学院、中国中医科学院、中国医学科学院与北京协和医学院是中国优秀的 DNA 条码科机构，陈士林和宋经元的文献较多，质量较高。在 SCI 收录的期 刊 中，中国没有 DNA 条形码研究领域的优秀期刊，在影响力较高 10 篇（LCS）的论文中有中国科学院昆明植物研究所李德珠、高连明等中国植物条形码研究团队（China Plant BOL Group）合作的论文《Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants》(Li et al, 2011)。

虽然我国开展 DNA 条形码的研究起步较晚，特别是植物 DNA 条形码的研究到 2008 年才真正开始，但发展十分迅猛，发表了系列的相关文章，并参与了国际植物核心 DNA 条形码的评估与推荐，相信不久的将来我国在国际生命条形码研究中会发挥更大的作用。从我

国国务院发布的《中医药发展战略规划纲要（2016—2030 年）》，DNA 条形码在中药的应用将会不断增多，目前已经获得了不少的成果，《中药 DNA 条形码分子鉴定》、《中国药典中药材 DNA 条形码标准序列》、中药材 DNA 条形码鉴定系统（http://www.tcmbarcode.cn/china/），DNA 条形码将会为中药的鉴定，中药将走出国门，被世界接纳奠定基础，中国植物 DNA 条形码研究也是蒸蒸日上，中国科学院昆明植物研究所和中国科学院华南植物园共同完成的“中国植物 DNA 条形码研究”科技研究已经获得不错的成果。

植物 DNA 条形码 2017 年文献动态及发展前景，在前面的数据分析中没有涉及到 2017 年的文献数据，Web of Science 核心合集中对 2017 年植物 DNA 条形码进行文献检索分析，共获得 208 个记录，该领域研究蓬勃发展，文献量均高于其他年份，中国和美国发文量占了将近 50%（103 篇），主要的发文期刊和作者与前面的结果一致，主要文献为 *Plant DNA barcodes: applications today and in the future*（Kress，2017）。2017 年延续了前期的研究，并重点开展了以下几个方面研究：群落系统发育和物种编目，功能形状和物种的编目，物种之间的互相影响：识别未知的物种，物种的边界和生物多样性的发现，DNA 条形码的取证：商业产品、濒危物种、草药的原材料、民族植物学，物种和栖息地的保护。同时某些在特定的植物(*Codonopsis*)群中，仍然可以看到各种各样的 DNA 条形码标记测试的研究报告（Wang et al, 2017）。

自从十多年前 DNA 条码被引入到植物群落，DNA 条码在植物基础研究和应用研究中都得到了广泛的应用。植物系统学家尚未普遍接受 DNA 条码作为识别物种的核心工具的一个主要原因是，在大多数分类群中，没有一个标记能够完全区分物种。相比之下，生态学家更愿意寻找新的、独特的 DNA 条码应用，以解决他们的一些基本研究问题，因为总的来说，他们在由多种植物谱系组成的系统中工作，这些系统可以通过 DNA 条码位点的组合进行唯一的识别。展望未来，植物 DNA 条码将通过两种关键的方式为植物群落服务：1）建立一个更全面、通用的全球植物 DNA 条码库。实现为世界上所有植物提供 DNA 条码的通用库的目标仍然是遥远的未来，但是一旦实现，基础研究和应用研究都将受益匪浅。2）开发新的标记并采用新的测序技术，如 DNA 宏条形码（metabarcoding 或 eDNA）(Taberlet et al，2012)，随着测序技术的提高和费用的降低，DNA 条形码向基因组学方向发展 (Coissac et al，2016)。

在此次分析中，存在以下几方面的不足：（1）用 HistCite 软件以 LGS Count 为条件，选用 LCS 绘制引文编年图，有时候一些新的文章因为年代近，被引用次数暂时还不多，所以它们在图中节点中被引用次数不会很多，也存在一定的缺陷，有些近期的论文很重要，但是没有表现出来。（2）基于 Web of Sciences 数据库涵盖了多种世界范围内最有影响力的高质量英文期刊，但其侧重于收录母语为英语地区的期刊，来源期刊地域分布的不平衡性对分析结果有一定的影响。

**参考文献：**


BEGROW D, NILSSON H, UNTERSEHER M, et al, 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures[J]. Appl Microbiol Biotechnol, 87(1):99-108.

BHARATHI DG, 2013. Methods employed in the web of science and scopus databases to effect changes in the ranking of the journals [J]. Curr Sci India, 105(3): 300-308.

BHARGAVA M, SHARMA A, 2013. DNA barcoding in plants: Evolution and applications of in silico, approaches and resources[J]. Mol Phlyogenet Evol, 67(3):631.

BLAXTER M, 2004. The promise of DNA taxonomy[J]. Philosophical Transac Royal Soc B Biol Sci, 359(1444):669-679.

CHASE MW, COWAN RS, HOLLINGSWORTH PM, et al, 2007. A proposal for a standardised

protocol to barcode all land plants. Taxon, 56(2):295-299.

COISSAC E, Hollingsworth PM, Lavergne S, et al, 2016. From barcodes to genomes: extending the concept of DNA barcoding [J]. Mol Ecol, 25(7): 1423- 1428.

GARFIELD E, 2009. From the science of science to scientometrics visualizing the history of science with Histcite software[J]. J Informetr, 3(3): 173-179.

GU HT, WANG J, 2013. Forecast of the Development Tendency for Discipline Based on two Models——Take Bibliometrics as an Example[J].J Mod Inf, 33(2): 162-165.[顾洪涛, 王筠, 2013. 基于两种模型的学科发展趋势预测——以文献计量学为例 [J]. 现代情报, 33(2): 162-165.]

HEBERT PDN, CYWINSKA A, Ball S, 2003. Biological identification through DNA barcodes [J]. Proceedings Roy Soc B-Biol Sci, 270(1512):313-321.

HOLLINGSWORTH PM, 2011. Refining the DNA barcode for land plants [J]. P Natl Acad Sci USA, 108(49):19451.

KERSS WJ, WURDACK KJ, ZIMMER EA, et al, 2005. Use of DNA barcodes to identify flowering plants [J]. Proc Natl Acad Sci USA, 102(23):8369-8374.

KRESS WJ, 2017. Plant DNA barcodes: applications today and in the future [J]. J Syst Evol, 55(4):1-17.

HOLLINGSWORTH ML, ANDRA A, FORREST LL, et al, 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species‐level sampling in three divergent groups of land plants[J]. Mol Ecol Resourc, 9(2):439.

KIM HM, OH SH, BHANDARI GS, et al, 2014. DNA barcoding of Orchidaceae in Korea [J]. Mol Ecol Resourc, 14(3):499-507.

LAHAYE R, BANK MVD, Bogarin D, et al, 2008. DNA barcoding the floras of biodiversity hotspots [J]. Proc Natl Acad Sci USA, 105(8): 2923-2928.

LI DZ, GAO LM, LI HT, et al, 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants [J]. Proc Natl Acad Sci USA, 108(49):19641-6.

LIU J, MOLLER M, GAO L M, et al, 2011. DNA barcoding for the discrimination of Eurasian yews (Taxus L. Taxaceae) and the discovery of cryptic species [J]. Mol Ecol Resourc, 11(1):89.

LU S, DING X, MA Y, et al, 2010. Confirming the Genetic Identity of Dendrobium fimbriatum Using an Amplification Refractory Mutation System (ARMS) [J]. Plant Mol Biol Rep, 28(4):712-716.

MOHAMMED AB, MOHD SF, SHAMSIR MS, et al, 2017. Review: dna barcoding and chromatography fingerprints for the authentication of botanicals in herbal medicinal products[J] Evid-Based Compl Alt, (1):1-28.

PEGARD A, MIGUEL C, VALENTINI A, et al, 2009. Universal DNA-Based Methods for Assessing the Diet of Grazing Livestock and Wildlife from Feces[J]. J Agric Food Chem, 57(13):5700.

REN BQ, XIANG XG, CHEN ZD, 2010. Species identification of Alnus (Betulaceae) using nrDNA and cpDNA genetic markers [J]. Mol Ecol Resourc, 10(4):594.

SHADBOLT N, HALL W, HENDLER JA, et al, 2013. Introduction: Web of Science: a new frontier [J]. Philosoph Transac Royal Soc A: Mathem Physical & Engineer Sci, 371(371):1-6.

SUN XH, LU WR, 2012. Research Trends of Rice Science Based on Web of Science [J]. Chin J

Rice Sci, 26(5):606-614.[孙秀焕, 路文如, 2012. 基于 Web of Science 的水稻研究态势分析[J]. 中国水稻科学, 26(5):606-614.]

TABERLET P, COISSAC E, POMPANON F, et al, 2012.Towards next generation biodiversity assessment using DNA metabarcoding [J]. Mol Ecol, 21(8): 2045–2050.

TEZCAN M, VLACHONASIS K, AKI C, 2010. DNA barcoding study on Sideritis trojana Bornm. an endemic medicinal plant of Ida Mountain, Turkey[J]. Fresen Environ Bull, 19(7):1352-1355.

VALENTINI A, POMPANON F, TABERKET P, 2009. DNA barcoding for ecologists [J]. Trend Ecol Evol, 24(2), 110-117.

WANG DY, WANG Q, WANG YL, et al, 2017. Evaluation of DNA barcodes in *Codonopsis* (Campanulaceae) and in some large angiosperm plant genera [J]. PLoS ONE 12: e0170286.

WU RM, GONG YF, 2010. A comparison on Co-word analysis and Co-citation analysis[J]. Inf Doc Serv, (1): 26-29. [伍若梅和孔悦凡, 2010. 共词分析与共引分析方法的比较研究[J]. 情报资料工作, (1): 26-29.]

YANG H, WANG XP, GAN WZ, et al, 2013. Analysis on literature development trend of international tea polyphenols research based on the Web of Science[J].J Tea Sci, (6):541-549. [杨华, 王小萍, 干文芝, 等, 2013. 基于 Web of Science 的国际茶多酚类研究文献发展态势分析[J]. 茶叶科学,(6):541-549.]

ZHAO RL, XU LM, 2010. The knowledge map of the evolution and research frontiers of the bibliometrics[J]. J Lib Sci China, 36(5):60-68.[赵蓉英, 许丽, 2010. 文献计量学发展演进与研究前沿的知识图谱探析[J]. 中国图书馆学报, 36(5):60-68.]